DOCUMENT RESUME

ED 182 323 TM 010 005

AUTHOR Hunter, John E.; Gerbing, David W.
TITLE Unidimensional Measurement and Confirmatory Factor Analysis. Occasional Paper No. 20.
INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY National Inst. of Education (DHEW) Washington, D.C. Teaching Div.
PUB DATE May 79
CONTRACT 400-76-0073
NOTE 75p.
AVAILABLE FROM Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, MI 48824 ($3.00)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cluster Analysis; Cluster Grouping; Cognitive Measurement; *Factor Analysis; Item Analysis; Mathematical Models; *Measurement; Oblique Rotation; Questionnaires; Test Construction; Test Reliability
IDENTIFIERS *Confirmatory Factor Analysis; Latent Trait Theory; *Test Theory

ABSTRACT

Confirmatory factor analysis is presented as providing appropriate techniques for the analysis and evaluation of questionnaires and tests if the content of the measure can be identified as consisting of groups of items, with each group measuring only a single trait. This approach is contrasted with latent trait theory which assumes (and does not test) the assumption of unidimensionality. The first step in the analysis is the partitioning of the items into distinct scales or clusters, preferably on the basis of an a priori analysis of item content, although blind exploratory statistical methods may be substituted. The parameters of the model are estimated by oblique multiple groups factor analysis and include the correlations between items and factors, the correlations between the factors, and the communality of each item. This analysis will confirm, or disconfirm, the measurement model specified by the user. Chi square tests are provided for unidimensionality of clusters and parallelism. Appendices contain proofs and computer programs. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED182323

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. E. Hunter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Institute for Research on Teaching

College of Education — Michigan State University

TM010 005

Occasional Paper No. 20

UNIDIMENSIONAL MEASUREMENT AND CONFIRMATORY FACTOR ANALYSIS

John E. Hunter and David W. Gerbing

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

May 1979

This work is sponsored in part by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Teaching Division of the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-76-0073)

Institute for Research on Teaching

Teachers' thoughts and decisions are the focus of studies currently under way at Michigan State University's Institute for Research on Teaching (IRT). The IRT, founded in April 1976 with a $3.6 million grant from the National Institute of Education, has major projects investigating teacher decision-making, including studies of reading diagnosis and remediation, classroom management strategies, instruction in the areas of language arts, reading, and mathematics, teacher education, teacher planning, effects of external pressures on teachers' decision, and teachers' perceptions of student affect. Researchers from many different disciplines cooperate in IRT research. In addition, public school teachers work at IRT as half-time collaborators in research, helping to design and plan studies, collect data, and analyze results. The Institute publishes research reports, conference proceedings, occasional papers, and a free quarterly newsletter for practitioners. For more information or to be placed on the IRT mailing list please write to: The IRT Editor, 252 Erickson, MSU, East Lansing, Michigan 48824.

Director: Lee S. Shulman

Associate Director: Judith E. Lanier

Editorial Staff:

Lawrence W. Lezotte, coordinator of Communications/Dissemination
Linda Shalaway, IRT editor
Janet Flegg, assistant editor

Contents

Preface iii

Introduction 1

Questionnaire Measures 2

Manifest Variables and Latent Variables 4

Clusters of Items 5

Construction of the Model 8

Analysis of Content 9

"Blind" Analysis 11

Estimating the Parameters of the Model 12

Evaluation of the Model 17

A First Statistical Test for Unidimensionality: Internal Consistency 18

A Second Test for Unidimensionality: Parallelism 26

Homogeneity of Content 33

Reliability 33

Revisions of the Model 34

References 37

Figures

1. A Unidimensional Measurement Model 19

2. The Relation Between a Cluster of Items and Another Group Factor 27

3. A Two Factor Multiple Groups Measurement Model . . . 29

Tables

1. The Output of a Multiple Groups Analysis 15

2. Unidimensional Correlation Matrices without Sampling Error, i.e., Matrices of Population Correlations 22

3. A Unidimensional Matrix with Sampling Error 24

4. A Non-unidimensional Correlation Matrix 25

5. An Example of Perfect Parallelism 30

6. An Example of Item-cluster Correlations showing Departure from Parallelism 31

7. An Example of a Matrix of Similarity Coefficients 32

Appendices

1. The Use of PACKAGE and LISREL IV for Confirmatory Factor Analysis 40

2. The Computation of a PACKAGE "Oblique Multiple Groups Factor Analysis" 52

3. The Meaning of Communalities 56

4. The General and Unidimensional Factor Models 58

5. Proof of the Product Rules for Internal and External Consistency 61

Preface

Much of the current research on teaching demands methods of analysis that have not been typically part of the armamentarium of researchers trained in traditional experimental analysis and design. The laboratory and quasi-laboratory contexts are giving way to *in vivo* contexts of real classrooms, pupils, and teachers. Instruments are being developed which, with constructs such as locus of control and self-regulations, are being used in place of or in addition to established achievement and personality tests. The dimensionality and reliability of these new instruments are often unknown.

All of these ongoing changes require the development of new analytical tools for parameter estimation and hypothesis testing. The paper which Hunter and Gerbing present here outlines these changing conditions of research and details implications of these changes for analysis. Their work describes the theory and the use of two computer programs for the use of some of these analytic tools for determining the adequacy of the "measurement model." The paper's appendices contain examples of the use of these programs.

Joe L. Byers

Unidimensional Measurement and Confirmatory Factor Analysis

John E. Hunter and David W. Gerbing[1]

Introduction

The goal of most scientific research is the investigation of the causal relations among a set of variables. Within the social sciences, especially psychology and education, the variables of interest are often traits such as abilities, attitudes, aptitudes, opinions, or sentiments. Although many studies do not go beyond describing the correlations between traits, more information may be obtained by investigating the causal relations of the variables. A network of causal relations between the variables is called a "causal model" or sometimes a "structural equation model." In pictorial form the model is called a "path diagram." Once a causal model has been hypothesized, the parameters can be estimated and the model can be tested and revised using path analysis. Blalock (1971), Duncan (1975), and Heise (1975) provide introductory discussions of this process.

The researcher must first measure the traits before the relations between them can be analyzed. The problem is that observed measurements are never perfect and are often only indirect estimates of the intended traits. Thus, observed values should not be used as measurements unless there is an accounting of the nature and degree of error in the observations. We will call this accounting a "measurement model." Most current measurement

---

[1]John E. Hunter is a professor of psychology at MSU. David W. Gerbing is an MSU doctoral student majoring in psychology.

problems can be cast in the language of "cluster analysis" (Tryon, 1939) or "Spearman factor analysis" (Spearman, 1904) or "congeneric tests" (Jöreskog, 1971, 1978) and that will be the approach outlined below.

In summary, most correlational work falls into two phases: measurement and theory testing. Since each of these phases can be mathematized using existing techniques, the full research process can be conceptualized as the construction and analysis of two models: a measurement model and a causal model. The construction and analysis of the measurement model is the subject of this paper.

There are two levels to most empirical studies, a molar and a molecular level. At the molar level, the investigator is concerned with the relations between theoretically defined variables. At the molecular level, the investigator has a large number of observations or observed responses from which he or she wants to infer the values of those theoretical variables for particular people at a particular point in time. The relationship between the molar and molecular level is the content of theories of measurement in the social sciences.

Questionnaire Measures

For simplicity, the general issues will be couched in terms of questionnaire data, though the principles apply equally well to experimental observations, event data, etc., on either humans or rats. For a questionnaire, the molar level might be represented by traits; the molecular level is given by the person's responses to the individual items that define the questionnaire. Several different kinds of traits may be measured by a variety of items.

Educational psychologists often use achievement and aptitude tests. An achievement test uses true-false or multiple choice items to assess specific knowledge such as:

> Identify the author of the statement, 'The vice-presidency isn't worth a pitcher of warm spit.'
>
> (a) Walter Mondale (b) Thomas Jefferson (c) John Garner

Aptitude tests are intended to serve as measures of ability to learn rather than knowledge mastered. These tests ask people to solve problems and the solutions are coded as "right" or "wrong."

Personality scales and certain attitude scales ask people to respond to each item on a scale ranging from "strongly disagree" to "strongly agree." Personality scales present self-descriptions such as:

I am a shy person

or descriptions of significant others such as:

My mother never fully approved of us.

Attitude scales ask for opinions on statements about policies or politicians such as:

Capital punishment should be outlawed.

Attitude scales such as the semantic differential ask people to rate objects on bipolar scales such as:

My mother-in-law: worthless _ _ _ _ _ valuable

or use behavioral rating scales such as:

My mother-in-law: is docile as a deer _ _ _ _ _ could make Jaws into a pizza.

In industrial psychology such behavioral scales may have a label or "anchor" at each point of the rating scale as in:

drops and breaks things	misplaces things	puts things away	checks things for maintenance

Checklists, matching items, and rank orders have also been used, but they are poor procedures since the responses are *not* independent measures.

Manifest Variables and Latent Variables

The relationship between trait and item is usually quite imperfect. For example, consider the trait "Knowledge of U.S. History" and the item "Who said, 'The vice presidency isn't worth a pitcher of warm spit'?" People who know a great deal of American history are more likely to get this item right than persons who know little history. However, at best this one item would merely place people into two categories, "high" and "low", and it could not register the many gradations of knowledge that we know to exist.

Another problem is that the classification of "right" or "wrong" on any one item is not likely to be a perfect indication of whether a person is "high" or "low" in knowledge of history. It is likely that some who are well versed in American history will have only a cursory knowledge of politics in the forties and hence will not recognize the statement as that of John Nance Garner after having served as vice president to Franklin Roosevelt. On the other hand, there are probably some people of little knowledge who heard the quotation on an election special and found it so quaint that it stuck. Furthermore, there may not even be a perfect relationship between knowledge of the item and generation of the correct response. Many's the time a person has a fact on the tip of the tongue but finds it

temporarily "blocked", i.e., a person who knows the answer might temporarily block on recall. On the other hand, if the item is presented in multiple choice form, then a person with no knowledge at all would still have a 20-25 percent chance of getting the right answer by guessing. We will refer to differences between trait and item as "error." In the preceding example two kinds of measurement error are illustrated: randomness in a particular response process and idiosyncracy or specificity of knowledge or relevant experience.

Thus the distinction is drawn between the manifest variables or items and the latent variables or traits. The traits are the variables of interest and the purpose of the manifest variables is to measure the corresponding traits. Other common words for traits are molar variables, true scores, domain scores, universe scores, constructs, unobserved variables, underlying variables, latent variables, theoretical variables, and factors. Manifest variables are also referred to as items, molecular variables, observed variables, fallible measures, indicators, proxy variables, and overt variables.

Clusters of Items

Researchers are well aware of error in molecular responses and they have a standard answer to that problem: multiple observations. For example, a history test might consist of 100 items or an attitude scale might have 10 items. The key idea is this: If several items measure the same trait, then the influence of trait on response is common to all the responses while the processes which create errors vary randomly from one item to the next. Thus the common threat running through the set of responses should be much more indicative of the trait than is any single

response. The most common way of finding the common thread through several responses is to add or average them. For example, if each item response to a history test is scored as 1 for a correct answer and 0 for an incorrect answer, then the sum of the item responses is the number of correct responses and the average of the 0's and 1's is the proportion of correct responses.

Sums or averages of responses are called tests, scales, indexes, operational definitions (though this is not the meaning intended in philosophy of science), to name just a few. The strategy we advocated is to organize the items on the questionnaire into clusters such that each cluster of items corresponds to a single underlying variable or trait. A cluster score is defined for each individual by summing or averaging across the individual's responses to the items in that cluster. These sums or averages provide a level of analysis which is intermediate to molar and molecular levels; the sums are more accurate estimates of a trait than are single item responses, but they are still not perfect. That is, although an individual's composite score is still only an imperfect estimate of the corresponding true score, it is a more reliable measure of the trait than an individual item.

The use of summation or averaging to combine observations is only optimal if the relationship between item responses and traits is approximately linear. If that relationship is highly non-linear, as in a Guttman (1944) scale, then other methods may be more efficient. This problem has been taken up within psychology under the rubric of "latent trait theory" (Lord, 1970; Lord & Novick, 1968; Urry, 1974). These theorists assume that the items given all measure the same trait and then proceed to the question: How do we estimate the trait score from the

given item score. That is, these methodologists ignore the question: "Do these items all measure the same trait? This more fundamental question is the focus of the present paper. For lack of space, we will ignore the question of non-linearity here and instead assume that the relationship between item and trait is linear.

Mathematically, any set of items can be defined as a cluster. And for any cluster, a score can be formed by summing over the responses to these items. However, if the cluster of items is to measure a single underlying trait, or construct, then the corresponding cluster score is interpretable only if the cluster is unidimensional. As Nunnally (1978) writes, "Items within a measure are useful only to the extent that they share a common core--the attribute which is to be measured" (p. 274). Other common terms used for unidimensional scales are homogeneous scales, unifactor scales and congeneric tests.

The relationship between the molar and molecular variables--the traits and the items--is formally specified by the measurement model. In this paper the measurement model specifies that all the items on a questionnaire should be partitioned into unidimensional clusters. The process of measuring traits by constructing unidimensional subscales is called "cluster analysis" (Tryon, 1939), though in recent years the meaning of this phrase has been refined. According to recent terminology, a "cluster analysis" refers to the grouping or clustering of items according to any one of the numerous clustering algorithms. The empirical evaluation of the unidimensionality of a cluster is a specific example of a "confirmatory factor analysis" after Jöreskog (1966, 1969).

The following procedure is used to test a unidimensional measurement model. First, construct a specific model by partitioning a set of items into subscales. Second, collect the data and then, using a program such as PACKAGE (Hunter & Cohen, 1969), compute the inter-item correlation matrix. Third again using a program like PACKAGE, estimate the parameters of the model with a confirmatory factor analysis of the cluster solution, i.e., an oblique multiple groups factor analysis (Harmon, 1976; Nunnally, 1978) with communalities. Fourth, test the fit of the model to the data by the three criteria of unidimensionality: homogeneity of content, internal consistency, and parallelism. Homogeneity of content is based on the meaning of the items. The dimensionality of each cluster is statistically evaluated by testing the correlations between the items in each cluster for internal consistency and by testing the estimated correlations between items and traits for parallelism. If the fit is not satisfactory, the model is modified by rearranging and deleting items and repeating the process until a set of unidimensional scales are obtained.

Construction of the Model

The first step in the analysis is the partitioning of the items into distinct scales, i.e., each item is placed in only one cluster. The initial set of clusters may be defined in two ways: on the basis of an a priori analysis of item content or on the basis of the inter-item correlations. Purely statistical procedures for generating clusters are called either "exploratory" or "blind."

Analysis of Content

During all phases of the analysis the researcher should have a theory of how a set of items can be written for each trait, i.e., a theory relating item responses to traits by a clustering of the items. The most important criterion for clustering at each phase of the analysis is item content or the meaning of the items.

The initial measurement model should be constructed before the data is collected. That is, there should be a rationally constructed a priori measurement model which is ready for parameter estimation and evaluation before the first computer run. At this point, the only basis for the writing and partitioning of the items is item content. Usually, there is an indefinitely large set of items which can be used as alternate indicators or measures of each trait. Writing items can be thought of as generating a sample from that item domain (Tryon, 1959) or universe of content (Guttman, 1944). Each item should be concise and unambiguous and the meaning of all the items in a cluster should be similar. The items should be worded so that they will discriminate between the individuals in the sample. If everyone agrees or everyone disagrees with an item, the item would have no variance and thus would not correlate with anything else. An excellent treatment of problems in item writing is given in Ebel (1972).

Items should not be mixed together in the same cluster unless the items share a specific and common meaning. A frequent mistake is to lump items together which are only superficially related. After a tentative cluster is written, set it aside and come back to it later. Unless you have had lots of practice, a careful reading of the items will probably reveal that further subclustering is possible. Simply on the basis of

meaning alone a purification of the scales at this stage often forces the investigator to be very explicit regarding the exact nature of variables he or she is interested in. The result is that the theory relating the variables is often refined and sharpened even before data collection begins. Moreover, the time to beef up the clusters with three or fewer items is before the data have been collected.

For example, suppose a researcher is interested in evaluating the perceptions of the parents of elementary school children to an innovative educational program. One of the variables the researcher is interested in measuring is the parents' perception of how the program has affected their children's development during the years the program has been instituted. The following five Likert scale attitude items were written to measure perceived development:

1. The program helps children to read better than the old program.
2. Most children who participated in the new program have made many more friends than children in the old program.
3. Children in the new program learn more.
4. Children in the new program enjoy doing homework.
5. The new program encourages children to work together.

The problem is that Items 1, 3, and 4 involve cognitive development the content of the items allows the researcher to postulate that two potentially distinct traits are being measured by these items. The items should be explicitly partitioned into two clusters and more items should be written so that each cluster contains, for example, five or six items. If only two or three items were retained on each scale, the measurement of the underlying traits would probably be very unreliable.

"Blind" Analysis

An alternate technique for defining cluster membership is to let a "blind" set of computational procedures define the clusters after the data have been collected. Instead of specifying a measurement model in advance, the investigator lets the computer uncover a structure purely on the basis of the correlational structure of the items, i.e., without consideration of meaning.

One popular blind algorithm for generating clusters, accomplished by the PACKAGE subprogram FACTOR is an exploratory factor analysis of the inter-item correlation matrix -- a principal axis factor analysis with communalities followed by a varimax rotation. For each factor a corresponding cluster is defined. The items are assigned to the clusters on the basis of their factor loadings, i.e., the correlations of the items with the factors. Each cluster is made up of those items whose highest loadings are on the corresponding varimax factor. In fact, many researchers stop at this point. They define a set of scales based on an exploratory factor analysis, and then never formally evaluate the scale for unidimensionality with a confirmatory factor analysis.

Many other clustering algorithms have been programmed. The user may use any of these blind techniques and then submit the obtained cluster structure to PACKAGE for evaluation of unidimensionality and reliability. The restriction is that the chosen clustering method should group the items into mutually exclusive subsets.

Another possibility is the use of the PACKAGE subprogram ORDER (Hunter, 1973). This program can be used to order items in terms of either

their correlation coefficients or similarity coefficients (see page 31). Items are ordered so that those which are highly correlated (or similar in pattern) are adjacent in the list and hence the high correlations (or similarity coefficients) are near the diagonal. Partitioning the ordered item list and the reordered correlation matrix can be done by looking for "break points" in either the content of the items or in the size of the correlations.

Estimating the Parameters of the Model

The parameters of a measurement model are estimated by a "factor analysis." For the unidimensional measurement models discussed in this paper, a particular factor analytic procedure is used called "oblique multiple groups factor analysis" (Gorsuch, 1974; Harmon, 1976; Nunnally, 1978). A "group factor" is simply the trait measured by a given cluster or group of items. The word "oblique" means that the traits may be correlated with one another. The word "multiple" refers to the existence of several clusters of items in the analysis. The input to a multiple groups analysis is the matrix of sample correlations between each pair of items on the questionnaire and the designated partitioning of the items into clusters. The output of a multiple groups analysis includes the parameter estimates which provide information for use in the evaluation of the unidimensionality of a set of clusters. The parameters of the measurement model always include the factor loadings, the correlations of the items and factors, and the correlations between the factors. A third kind of parameter which will be explained later is the communality of each item. Communalities are usually included in the analysis, but they are calculated at the user's option.

An important characteristic of a multiple groups analysis is that

it is confirmatory. The measurement model is specified by the user; the number of clusters (or factors) and the composition of the clusters is specified in advance. The multiple groups analysis estimates parameters on the basis of the given model. The model is then tested for goodness of fit, i.e., the unidimensionality of the clusters is evaluated.

An oblique multiple groups analysis can be performed with at least three computer programs: Tryon's BCTRY (Tryon & Bailey, 1970); Jöreskog's LISREL (Jöreskog & Sörbom, 1978); and Hunter's PACKAGE (Hunter & Cohen, 1969). BCTRY and LISREL are more general programs than PACKAGE and are therefore more difficult and more expensive to use. Thus any reference to computation in this paper will usually be in the language of PACKAGE. Examples of the use of the PACKAGE subprograms discussed in the paper are listed in Appendix 1 along with the input for a multiple groups analysis using LISREL. The computational details of a PACKAGE multiple groups analysis are presented in Appendix 2.

A preliminary step in the analysis is the computation of the correlation matrix from a standard person-by-variable data matrix. The PACKAGE subprogram CORR will perform this task with or without missing data. If the correlations have been computed previously either by PACKAGE or by a program such as SPSS, the matrix can be read into memory with the PACKAGE subprogram READR.

An oblique multiple groups analysis can be performed with the PACKAGE subprogram MGRP. The usual use of MGRP is to perform a multiple groups analysis on the user-defined cluster structure imposed on the inter-item correlation matrix. However, MGRP may also be accessed indirectly with the subprogram FACTOR. The FACTOR subprogram will perform a blind analysis, i.e., after defining clusters from a classical factor analysis,

the "blind" clusters are automatically submitted to MGRP for a multiple groups analysis. That is, FACTOR performs both an exploratory *and* a confirmatory factor analysis.

The output of a multiple groups analysis is one large correlation matrix which may be partitioned into the three following correlation matrices: (a) the inter-item correlations arranged by cluster, (b) the correlations between the items and the factors, and (c) the correlations between the factors (the correlations to be used in the causal model). The general form of this matrix and a specific example from PACKAGE are presented in Table 1. A PACKAGE convention is to number the group factors from 501 while restricting item numbers to 1-499. For example, $r_{7,502}$ is the correlation between item 7 and the second group factor.

Any set of factor analytic computations can be done either on the given correlation matrix with ones in the diagonal or with the diagonal ones replaced by smaller numbers called "communalities." The choice of the diagonal elements is not arbitrary, but determines the definition of the factors to be defined in the analysis. The technical reasons for this are given in Appendix 3.

If ones are placed in the diagonal, then the item-factor correlations are the correlations between the items and the observed cluster score rather than its underlying trait. Each cluster is defined by a set of items. An individual's observed score on that cluster is simply the average or sum of the individual's score on each of the relevant items. The correlation of an item with its own cluster score is the item-total correlation, and is greatly inflated by a spurious common error term. The factor-factor correlations are the correlations between observed scales and are attenuated or reduced by measurement error. However, these

Table 1: The output of a multiple groups analysis.

Correlations between items (usually with communalities)	Correlations between items and factors
Correlations between items and factors	Correlations between factors

a) The general form of the output of a multiple groups analysis performed by PACKAGE.

	66	58	67	25	60	87	79	103	501	502
66	.49	.48	.49	.40	.41	.08	.13	.14	.70	.20
58	.48	.46	.44	.35	.43	.14	.09	.18	.68	.24
67	.49	.44	.46	.47	.35	.04	.13	.14	.68	.17
25	.40	.35	.47	.36	.40	.06	.05	.14	.60	.14
60	.41	.43	.35	.40	.35	.09	.00	.13	.59	.12
87	.08	.14	.04	.06	.09	.42	.42	.29	.13	.65
79	.13	.09	.13	.05	.00	.42	.42	.29	.13	.64
103	.14	.18	.14	.14	.13	.29	.29	.21	.22	.45
501	.70	.68	.68	.60	.59	.13	.13	.22	1.00	.27
502	.20	.24	.17	.14	.12	.65	.64	.45	.27	1.00

b) A specific output of a PACKAGE multiple groups analysis with communalities. Items 66, 58, 67, 25, and 60 define the first cluster and items 87, 79, and 103 compose the second cluster.

correlations are simple Pearson correlations and have the usual statistical distribution of sampling error.

If communalities are placed in the diagonal of the original correlation matrix, then the parameters of the model are based on cluster true scores. The item-factor correlations are the estimated correlations between items and cluster true scores, and the factor-factor correlations are the estimated correlations between cluster true scores. The cluster true score (the psychometric term for the score on the underlying traits) is that score which would be obtained if the underlying variable were measured without error. This option is generally preferred because (a) correlations of items and true scores are corrected for the error involved in measuring the variables with only a small number of items, and (b) the correlation of an item with its own cluster is not artificially inflated by correlating an item with an entity of which it is a substantial part, as is the case with an item-total correlation. The cluster true score is the average score of the entire indefinitely large domain of items, instead of only the few items appearing on the questionnaire. The use of communalities implicitly corrects for attenuation and hence eliminates the effect of error of measurement from the estimated correlations between items and factors or between factors and factors. However, correction for attenuation does not eliminate sampling error. In fact, the sampling error in these corrected correlations is larger than would be predicted by conventional formulas (Hunter, 1977). And the accuracy of these correlation estimates depends on the fit of the model. If a cluster is not unidimensional, then the correlations for the corresponding cluster true score are suspect.

In terms of computation, if the use of communalities is desired, as

will normally be the case, then the PACKAGE communalities option, COMM, should be used. This option may be used with either the MGRP or FACTOR subprograms.

Evaluation of the Model

Regardless of whether the clusters were formed from a theory or through a "blind" data analysis, the actual analysis of unidimensionality consists of evaluating each of the clusters according to these three criteria. The items in a unidimensional cluster are (a) internally consistent, (b) parallel or externally consistent, and (c) share a similar meaning. Failure to meet any one of these criteria is grounds for dropping an item from a cluster; each of these criteria are necessary but none are sufficient properties for an item to be accepted as an alternate indicator of the underlying trait.

Although content is the most important criterion, the statistical criteria of internal and external consistency are also necessary for evaluation of the model. The analysis begins with a content based theory of how the items should be partitioned and then proceeds to test that theory empirically. The empirical analysis of responses to the items may provide evidence which supports alternative theories. The meaning of some of the items as perceived by the researcher may be different from the meaning of these items as perceived by the respondents. As Loevinger (1948, p. 512) writes in the context of item domains, "there are many areas of psychological testing where the investigator's judgment is far from infallible." The researcher may disregard "subtle" features of content which are perceived as important by the respondents. For example, an item writer focused on one idea may not realize that a certain word is ambiguous, that it is open to a different interpretation. The people

responding to the item have no such bias and may thus generate answers that are irrelevant to the assessment of the desired trait. These item failures will be detected by the statistical analysis since these items will not be "parallel" to the other items in their cluster.

The number of items in each cluster cannot be determined on the basis of content alone. A prudent item writer provides for the possibility of failure with advance planning. Any important idea should be represented by at least three items. Thus, if one item fails, it can be detected as the "odd" item in the statistical analysis. The maximum number of items needed depends on the statistical quality of the items and on the number of persons in the study. The elimination of error of measurement using correction for attenuation or communalities increases the extent of sampling error. Studies with small sample sizes should use fewer but longer scales in order to minimize the influence of measurement error and hence the difference between the corrected and uncorrected correlation coefficients.

A measurement model may fail for reasons more profound than a few poorly written items; there may be very different dimensions determining the responses than those imagined by the investigator. For example, some of the ostensible arithmetic items may actually measure reading ability; a subset of "tricky" items may actually measure test anxiety. That is, there may be an entirely different way of clustering the items which is more appropriate to the dimensions which actually determine the responses. This could only be determined by empirical analysis (for example, exploratory factor analysis as done by the PACKAGE subprogram FACTOR).

A First Statistical Test for Unidimensionality: Internal Consistency

The causal relations between the items of a unidimensional cluster are

determined by their causal relationship to the underlying trait. This relationship can be shown in its most general form in a causal diagram such as that shown in Figure 1, where T is the score on the trait, X_i is the response to item i, and e_i is the net effect of extraneous causal factors for item response X_i.

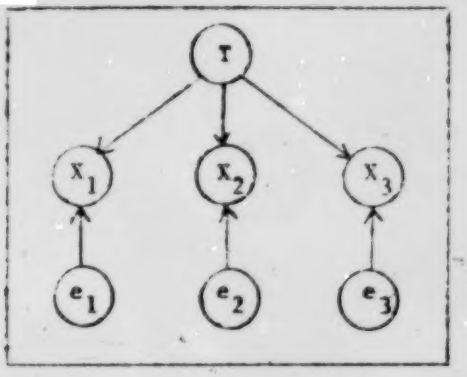

Figure 1: A unidimensional measurement model.

If the relationship between each item and the trait is linear, then this diagram can be translated into linear regression equations; for example:

$$X_1 = \beta_1 T + e_1$$
$$X_2 = \beta_2 T + e_2 \qquad \text{or} \qquad X_i = \beta_i T + e_i$$
$$X_3 = \beta_3 T + e_3$$

where β_i is the slope of item response X_i onto T, and e_i is the error in the regression of X_i on T. By definition, the errors of regression are uncorrelated with T, i.e., $r_{Te_i} = 0$. However, the extent of the correlation between errors for different items constitutes the crux of the assumption of unidimensionality. If the only significant causal factor determining each item response is the given underlying trait, then the

errors e_i should be uncorrelated with each other. In a linear model, this is the definition of unidimensionality. The relationship between the unidimensional factor model and the general factor model is presented in Appendix 4.

The problem for the researcher is to determine if the items which define a cluster satisfy this definition of unidimensionality. One strategy is to examine the observed correlations of the items in a cluster with each other and with items or other variables outside the cluster. The assumption of unidimensionality implies certain patterns among those correlations. Checking for these patterns produces two formal tests of unidimensionality. The test for "internal consistency" uses the correlations between items in the same cluster. The test for "external consistency" or "parallelism" uses the correlations of the items in a cluster with variables outside of the cluster.

If all the items in a cluster measure the same factor, then the correlations between the items will satisfy a "product rule for internal consistency." If X_i and X_j are two items in the same unidimensional cluster and T is the cluster true score, then the correlation between the items should satisfy the product rule:

$$r_{X_i X_j} = r_{X_i T} r_{X_j T}$$

That is, the correlation between two items in the same cluster should be the product of their correlations with the underlying trait. There are then two steps to test for internal consistency: (a) estimate the parameters $r_{X_i T}$ from the data and (b) see if the product rule reproduces the inter-item correlations to within sampling error. The proof of this product rule is presented in Appendix 5.

In the preceding discussion the reference to two items implies that $i \neq j$ in the product rule above. What happens if the product rule is used with $i = j$?

$$r_{X_iX_i} = r_{X_iT}r_{X_iT}$$

$$= r^2_{X_iT}$$

The number "$r_{X_iX_i}$" is not the correlation between X_i and itself which would be 1.00, but this number is useful in its own right. This number $r_{X_iX_i} = r^2_{X_iT}$ is called the "reliability" or the "communality" of X_i. However, because of this product rule the notation "$r_{X_iX_i}$" is the traditional notation for the reliability even though it is ambiguous with the regular correlation notation of $r_{X_iX_i} = 1.00$. Since the communality $r_{X_iX_i}$ equals $r^2_{X_iT}$, the computation of communalities is tantamount to computation of the factor loadings r_{X_iT} and is often treated as such in factor analysis texts.

How can you tell if the correlations among the items in a cluster conform to the product rule for internal consistency? Some formal computational procedures are presented later, but the most useful guide is often a simple inspection of the correlation matrix. There are two basic patterns for unidimensional matrices. First, if all the items have equal quality, i.e., if all the items have the same correlation with the cluster true score, then any two items will have the same correlation (to within sampling error). In this case, the correlation matrix is said to be "flat." An example of a flat correlation matrix without sampling error is given in Table 2a. Each item in the four-item cluster correlates

Table 2: Unidimensional correlation matrices without sampling error, i.e., matrices of population correlations.

	1	2	3	4
1	.64	.64	.64	.64
2	.64	.64	.64	.64
3	.64	.64	.64	.64
4	.64	.64	.64	.64

communalities are equal

a) Unidimensional correlation matrix for items of equal quality, i.e., a "flat" unidimensional matrix.

Correlations Decrease

Correlations Decrease

	1	2	3	4
1	.81	.72	.63	.54
2	.72	.64	.56	.48
3	.63	.56	.49	.42
4	.54	.48	.42	.36

communalities ordered high to low

b) Unidimensional correlation matrix for items of unequal quality, i.e., a unidimensional matrix with a "strong-weak" gradient.

.80 with the cluster true score, i.e., $r_{X_iT} = .80$, so for each X_i and X_j: $r_{X_iX_j} = (.80)(.80) = .64$.

Second, if the items do not have uniform quality, then the correlation matrix can be arranged so as to show a "strong-weak gradient" (to within sampling error). If the items are ordered in terms of their true score correlation (or communality) from strong to weak, then the high correlations will be in the upper left-hand corner and the lowest correlations will be in the lower right-hand corner. An example of a correlation matrix with a perfect strong-weak gradient, i.e., no sampling error, is given in Table 2b. The items have the true score correlations

$$r_{X_1T} = .90,\ r_{X_2T} = .80,\ r_{X_3T} = .70, \text{ and } r_{X_4T} = .60,$$

which is all the information needed to generate the full matrix. For example, the correlation between items 2 and 3 is:

$$r_{X_2X_3} = r_{X_2T}r_{X_3T} = (.80)(.70) = .56$$

and the communality of the fourth item is:

$$r_{X_4X_4} = r^2_{X_4T} = (.60)^2 = .36$$

The preceding discussion assumed that all correlations are observed on the population; that is, that all correlations are observed without sampling error. Sampling error will produce chance deviations from the product rule. Had the correlation matrix in Table 2a been estimated from a sample of 200 persons, then the resulting sampling error would produce deviations such as those shown in Table 3. The strict ordering produced in correlations which follow the product rule breaks down, though the general pattern is still visible. Since communalities must be estimated from the observed correlations, the communalities are also subject to sampling error of about the same magnitude.

Table 3: A unidimensional matrix with sampling error.

	1	2	3	4
1	.95	.73	.61	.59
2	.73	.59	.50	.42
3	.61	.50	.38	.29
4	.59	.42	.29	.31

estimated communalities

From the existence of the approximate strong-weak gradient, the researcher should recognize that this matrix is unidimensional to within sampling error. Of course the more sampling error present in a study, the more difficult it becomes to distinguish between unidimensional and multidimensional matrices. Since the sampling error of a correlation coefficient is a function of the number of subjects used to compute the correlation, it is desirable to have at least 100 and preferably 200 or more people in the study.

Table 4a shows a non-unidimensional correlation matrix with communalities estimated assuming unidimensionality. This matrix has two stark departures from a strong-weak gradient. In the first row, the .70 is larger than either of the first two entries (.64 and .53). In the second row, the .59 at the end is actually the largest entry in that row. If the sample was very small (say 50 or less), then such departures might happen by chance. But for sample sizes greater than 100 the standard error of the correlation is less than .1 and departures that large would be suspect. The task now is to find an arrangement which is multidimensional but provides a better fit. Table 4b shows a rearrangement of the

Table 4: A non-unidimensional correlation matrix.

	1	2	3	4
1	.64	.53	.70	.34
2	.53	.56	.38	.59
3	.70	.38	.43	.29
4	.34	.59	.29	.31

Communalities estimated assuming unidimensionality

a) A non-unidimensional correlation matrix with variables listed by order of estimated communalities.

	1	2	3	4
1	.70	.70	.53	.34
2	.70	.70	.38	.29
3	.53	.38	.59	.59
4	.34	.29	.59	.59

Communalities estimated assuming a two-dimensional structure

b) A reordering of Table 4a to show the clustering of items 1 and 3 and items 2 and 4.

variables into two clusters: items 1, 3 versus items 2, 4. The high correlations .70 and .59 are now within cluster correlations while all the correlations between clusters are noticeably lower. If there is a cluster with only two items, then it cannot be tested for internal consistency. A correlation matrix with only one entry is "flat" by definition and hence cannot show a departure from a gradient. Fortunately a two-item cluster can still be tested for parallelism or external consistency.

A more rigorous test of the product rule for internal consistency is to partial out the trait variable using a program such as the PACKAGE subprogram PARTIAL. The formula for a partial correlation between X_i and X_j with T held constant is:

$$r_{X_iX_j \cdot T} = \frac{r_{X_iX_j} - r_{X_iT}r_{X_jT}}{\sqrt{1 - r_{X_iT}^2} \cdot \sqrt{1 - r_{X_jT}^2}}$$

The numerator of the above expression, $r_{X_iX_j} - r_{X_iT}r_{X_jT}$, is the "residual" form of the product rule, i.e., the difference between the obtained and reproduced correlations. This "residual" is therefore predicted to be zero. If the cluster is unidimensional, each such partial correlation should be zero to within sampling error. Thus, the computation of partial correlations represents a formal comparison between the correlations implied by the model using the available parameter estimates ($r_{X_iT}r_{X_jT}$) and the obtained sample correlations ($r_{X_iX_j}$).

A Second Test for Unidimensionality: Parallelism

The second statistical criterion for the evaluation of unidimensionality is parallelism. The criterion of internal consistency specifies how the

items comprising a unidimensional cluster should correlate with one another. The criterion of external consistency or parallelism specifies how these items should correlate with variables outside of the cluster. The general statement of parallelism is that the items in a unidimensional cluster have similar patterns of correlations with (a) items in other clusters, or (b) other traits. In fact, the correlations for items of the same quality should be equal (to within sampling error) across all other variables. Differences in the correlations for two items in the same cluster should be directly proportional to differences in the reliability of those items.

Consider three items X_1, X_2, and X_3 which are indicators of T and let U be some other latent variable or trait. If U and T are correlated, then each X is also correlated with U. These relationships can be illustrated by the following path diagram in which the curved double-headed arrow represents correlation with no specification of causality.

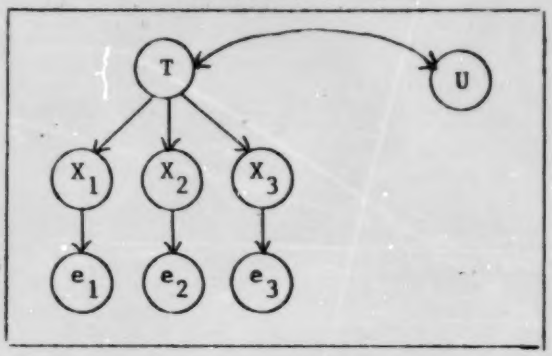

Figure 2: The relation between a cluster of items and another group factor.

The crucial fact in this diagram is that each X is correlated with U because T is correlated with U, and the size of r_{XU} depends directly on the size of r_{TU}. The formal statement of parallelism, called "the product

rule for external consistency", follows from the relationships specified in the diagram. This product rule is:

$$r_{XU} = r_{XT}r_{TU}$$

which is a generalization of the product rule for internal consistency. The proof of this product rule is presented in Appendix 5.

If this product rule for external consistency is applied to two items in a unidimensional cluster, X_i and X_j, then it follows from this product rule, according to Hunter (1973) and Tryon and Bailey (1970), that:

$$\frac{r_{X_iU}}{r_{X_jU}} = \frac{r_{X_iT}}{r_{X_jT}} = \lambda_{ij}$$

where λ_{ij} is a constant of proportionality for items X_i and X_j

For a unidimensional cluster, the ratio of the correlations between two items and a trait is equal to the ratio of their own respective true score correlations. Moreover, U could be any trait, so this same proportionality is maintained across all other variables. It is this form of the product rule which is used to evaluate a cluster for unidimensionality.

A second version of the product rule for external consistency is used to "explain" the correlation between two items in different clusters. This relation is illustrated in Figure 3, in which X_1, X_2, and X_3 are indicators of the trait T and Y_1, Y_2, and Y_3 are indicators of the trait U. As shown in Appendix 5, this model specifies that the correlation between items in different clusters is, for items X_i and Y_k:

$$r_{X_iY_k} = r_{X_iT}r_{TU}r_{UY_k}$$

If this version of the product rule for external consistency is applied to two items in a unidimensional cluster, X_i and X_j, then the

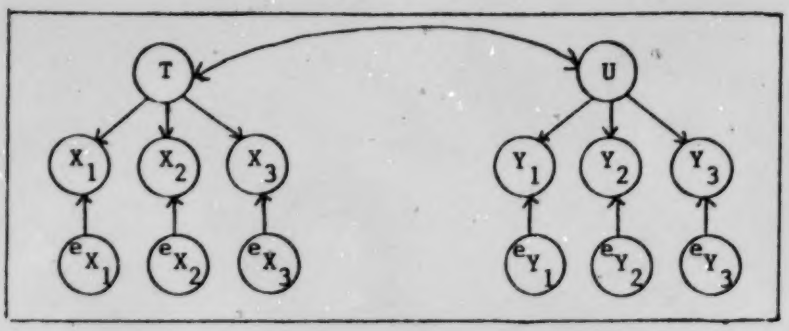

Figure 3: A two factor multiple groups measurement model.

same constant of proportionality results:

$$\frac{r_{X_i Y_k}}{r_{X_j Y_k}} = \frac{r_{X_i T}}{r_{X_j T}} = \lambda_{ij}$$

Thus, the parallelism product rule is applicable to items and to traits but it is usually applied to traits since the reliability of an item is lower than the trait it measures. As an example, consider a case of perfect parallelism between three items and four traits. The traits are named 501, 502, 503, and 504 in the notation of PACKAGE. Let the three items be part of a unidimensional cluster which is represented by variable 501. Thus, the correlations between the three items and variable 501 are the cluster true score correlations.

The example presented in Table 5 represents a case of perfect parallelism because each item has the same pattern of correlations with the four traits. The strongest correlations are those for X_1, the most reliable item, and the weakest correlations are those for X_3, the least reliable item. Moreover, each row of the correlation matrix (which represents the correlations of the items with one given trait) is exactly

Table 5: An example of perfect parallelism.

	.1	2	3
501	.80	.60	.40
502	.40	.30	.20
503	.72	.54	.36
504	.08	.06	.04

proportional or "parallel" to the first row, i.e., to the row of correlations between the items and their own true score. For example, $r_{X_3T} = .40$ and $r_{X_1T} = .80$, so the constant of proportionality is:

$$\frac{r_{X_3T}}{r_{X_1T}} = \frac{.40}{.80} = \frac{1}{2}$$

Each correlation of X_3 with 502, 503, and 504 is exactly half as large as the correlation of X_1 with the corresponding trait.

This example represents perfect parallelism because (a) the cluster containing X_1, X_2, and X_3 is perfectly unidimensional, and (b) there is no sampling error. Real data will contain sampling error so the investigator must decide if the deviations from perfect parallelism are due to sampling error only or are large enough to suggest multidimensionality.

One relatively extreme violation of parallelism is easy to detect. The product rule for external consistency implies that an item should correlate more with its own trait than with another trait. Unless two traits are highly correlated and there is more than a negligible amount

of sampling error, deviations from this principle are easy to detect. Consider, for example, the correlations of the four items with three factors shown in Table 6. The four items belong to the first cluster.

Table 6: An example of item-cluster correlations showing departure from parallelism.

	1	2	3	4
501	.80	.70	.70	.50
502	.40	.40	.20	.80
503	.30	.40	.30	.20

Item X_4 was originally placed in the first cluster, but it correlates .80 with the second cluster and only .50 with its own cluster. Thus, according to this analysis, items X_1, X_2, and X_3 should remain in the first cluster, but item X_4 should not. However, item X_4 might fit in with the second cluster if it meets the tests of content, internal consistency, and parallelism in that cluster.

A summary index is also available for specifying the parallelism of two items. Adaptation of the following formula, originally used as a measure of the similarities of two factors, is advocated by Tryon and Bailey (1970) and Hunter (1973).

$$\phi_{X_i X_j} = \frac{\sum_{k=1}^{n} r_{X_i Y_k} r_{X_j Y_k}}{\sqrt{\sum_{k=1}^{n} r^2_{X_i Y_k}} \sqrt{\sum_{k=1}^{n} r^2_{X_j Y_k}}}$$

Tryon and Bailey call the square of this index P^2 or the "index of proportionality", while Hunter refers to the index as a "similarity coefficient." If communalities are placed in the diagonal of the correlation matrix, the value of ϕ ranges from -1 to 1 with ϕ equaling 1 or -1 if the two items, X_i and X_j are perfectly proportional, i.e., perfectly parallel.

Similarity coefficients are computed by the PACKAGE subprogram SQRR. An excerpt of an output from SQRR for the two clusters of items presented in Table 1b is listed in Table 7. The full matrix contained over 100 items.

Table 7: An example of a matrix of similarity coefficients.

	66	58	67	25	60	87	79	103
66	1.00	.95	.94	.93	.94	.50	.48	.68
58	.95	1.00	.94	.91	.94	.54	.52	.72
67	.94	.94	1.00	.93	.93	.47	.47	.65
25	.93	.91	.93	1.00	.93	.45	.40	.64
60	.94	.94	.93	.93	1.00	.45	.40	.63
87	.50	.54	.47	.45	.45	1.00	.89	.84
79	.48	.52	.47	.40	.40	.89	1.00	.78
103	.68	.72	.65	.64	.63	.84	.78	1.00

The analysis of the correlations between the items defining a cluster is part of the analysis of internal consistency. The analysis of the similarity coefficients is part of the analysis of external consistency.

Homogeneity of Content

The most important criterion for evaluating unidimensionality is based on the meaning of the items -- not the statistical relations among the items. The criterion is that the items in a cluster should have similar meanings. If the items are indicators of the same construct, then the meaning of the indicators should fit together in a coherent, interpretable pattern. From the viewpoint of the social scientist, the cluster is useful only to the extent that the content of the cluster is unambiguously interpretable.

Moreover, content is also related to statistics. If cluster membership is determined solely on the basis of the sample correlations, sampling error will often lead to the formation of clusters which would probably not cross-validate in other studies. If clusters are formed only on the basis of statistics, then the resulting rearrangement of the correlations is likely to "capitalize on chance."

Reliability

If a cluster is unidimensional, then all the items measure the same trait. However, the amount of error of measurement in those items might be great or small. If the items are summed to form a cluster score, then the errors tend to cancel out so there is less error in a cluster score than in an item. However, if the individual items are very noisy and if the number of items in the cluster is small, then the amount of error in the cluster score could still be large. Thus, it is important to assess the quality of the cluster score in terms of the amount of error of measurement as well as in terms of unidimensionality. The measure of lack of error in a score is its "reliability."

Although the issues have often been confused in the literature, the dimensionality of the scale and the reliability of the cluster score are

independent concepts (Hunter, 1977; Green, Lissitz & Mulaik, 1977). One index of measurement error in a cluster score is coefficient alpha, which is also provided by the PACKAGE subprogram MGRP. The closer the value of coefficient alpha is to 1.00, the more reliable the measurement of the underlying variable. However, coefficient alpha provides an unbiased estimate of the reliability of the cluster score only if the scale is unidimensional. For multidimensional scales, alpha is an underestimate of the reliability of the composite score (Novick & Lewis, 1967). Thus, the value of alpha should be interpreted only after the cluster has been demonstrated to satisfy the three criteria of unidimensionality.

There is a version of coefficient alpha for multidimensional scales (Tryon, 1959; Rajaratnan, Cronbach & Gleser, 1965), but Moiser's (1943) theoretical formula for the reliability of a composite score is both more accurate and appropriate conceptually. Heise and Bohrnstedt (1970) use Moiser's formula with communalities as item reliabilities and call their estimate of composite reliability coefficient omega. However, Smith (1974) shows that omega also underestimates the reliability of a multidimensional composite, though not as much as alpha.

Revisions of the Model

Given the goals of unidimensionality and reliability, there will invariably be modifications of the original model. Items are deleted and reorganized, usually by dividing clusters into smaller groupings. Those unreliable items which are deleted from the original scales and not placed in another scale can be put in a "residual cluster." The unreliability of these items may be due to either one or both of the following two reasons: (1) An item may be poorly written so that its unreliability is due to random response error; each time the person

reads the item it is interpreted differently; or, (2) An item in the residual cluster may be a "one-item" cluster, i.e., a content domain represented by only one item. Such an item would be unidimensional by default, though its reliability would be unknown. However, there are methods of estimating reliability by using test-retest correlations in a longitudinal study as noted by Heise (1969).

A convenient strategy is to begin the analysis with both the a priori and blind measurement models, i.e., the use of both the MGRP and FACTOR subprograms contained in PACKAGE. Each analysis can be evaluated separately and the two solutions can be compared. If the content analysis is based on the correct dimensions, then the factor-analytically generated clusters will differ from those of the content analysis in being larger clusters in which some of the original content clusters may be merged. Exploratory factor analysis tends to blur distinctions between correlated clusters. Thus, if the content analysis is essentially correct, then the factor analysis usually generates a poorer cluster analysis. On the other hand, factor analysis is a good hedge against a content analysis which contains major errors. Factor analysis will usually find those dimensions which are completely outside the investigator's frame of reference.

The conceptual (MGRP) and the empirical (FACTOR) cluster solutions may be synthesized in the construction of a new model. The revised measurement model is then submitted to MGRP and another multiple groups output is obtained. At this point it is also useful to obtain the matrix of similarity coefficients from the SQRR subprogram since both correlation and similarity coefficients aid the evaluation of unidimensionality.

In each revision of the model, the between-item correlations and

similarity coefficients do not change, though rearranging the items by cluster membership aids evaluation. And the correlations between items and traits and between traits will change if the items which define the clusters change. Generally, the model is revised several times (especially if the sample size is small) until a decent fit is obtained. An additional check of the model is provided by the subprogram PARTIAL. When a cluster solution with communalities is obtained, the group factors can be partialed out of the correlation matrix. The partial correlations should be equal to zero within the limits of sampling error.

At this point, the reward for channeling a fair amount of effort into the construction of the scales is realized. The greater the initial effort, the more quickly a clean meaningful cluster solution emerges. Once each cluster is considered to reliably measure a single underlying trait, the measurement model is defined. That is, the traits have been defined in terms of the items and the large item correlation matrix has been reduced to the much smaller correlation matrix among the traits. Only this smaller correlation matrix between traits is subject to further analysis, i.e., the construction of a causal model.

References

Blalock, H. M., Jr. (Ed.). Causal models in the social sciences. Chicago: Aldine Publishing Company, 1971.

Duncan, O. D. Introduction to structural equation models. New York: Academic Press, Inc., 1975.

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1972.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. Limitations of coefficient alpha as an index of test unidimensionality. Educational and Psychological Measurement, 1977, 37, 827-838.

Gorsuch, R. L. Factor analysis. Philadelphia: W. B. Saunders Co., 1974.

Guttman, L. A basis for scaling qualitative data. American Sociological Review, 1944, 9, 139-150.

Harmon, H. H. Modern factor analysis (3rd Ed.). Chicago: University of Chicago Press, 1976.

Heise, D. R. Separating reliability and stability in test-retest correlation. American Sociological Review, 1969, 34, 93-101.

Heise, D. R. Causal analysis. New York: John Wiley and Sons, 1975.

Heise, D. R., & Bohrnstedt, G. W. Validity, invalidity and reliability. In E. F. Borgatta and G. W. Bohrnstedt (Eds.), Sociological Methodology. New York: Jossey-Bass, Inc., 1970.

Hunter, J. E. Methods of reordering the correlation matrix to facilitate visual inspection and preliminary cluster analysis. Journal of Educational Measurement, 1973, 10, 51-61.

Hunter, J. E. Cluster analysis: Reliability, construct validity, and the multiple indicators approach to measurement. Paper presented at a meeting of the U.S. Civil Service Commission, March 21, 1977.

Hunter, J. E., & Cohen, S. H. PACKAGE: A system of computer routines for the analysis of correlational data. Educational and Psychological Measurement, 1969, 29, 697-700.

Jöreskog, K. G. Testing a simple structure hypothesis in factor analysis. Psychometrika, 1966, 31, 165-178.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1969, 34, 183-202.

Jöreskog, K. G. Statistical analysis of sets of congeneric tests. Psychometrika, 1971, 36, 109-133.

Jöreskog, K. G. Structural analysis of covariance and correlation matrices. Psychometrika, 1978, 43, 443-477.

Jöreskog, K. G., & Sorbom, D. LISREL IV: Analysis of linear structural relationships by the method of maximum likelihood. Chicago: National Educational Resources, Inc., 1978.

Loevinger, J. A technique of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. Psychological Bulletin, 1948, 45, 507-529.

Lord, F. M. Item characteristic curves estimated without knowledge of their mathematical form--a confrontation of Birnbaum's logistic model. Psychometrika, 1970, 35, 43-50.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1968.

Moiser, C. I. On the reliability of a weighted composite. Psychometrika, 1943, 8, 161-168.

Novick, M. R., & Lewis, C. Coefficient alpha and the reliability of composite measurements. Psychometrika, 1967, 32, 1-13.

Nunnally, J. C. Psychometric theory (2nd Ed.). New York: McGraw-Hill Book Company, 1978.

Rajaratnam, N., Cronbach, L. J., & Gleser, G. C. Generalizability of stratified-parallel tests. Psychometrika, 1965, 30, 39-56.

Smith, K. W. On estimating the reliability of composite indexes through factor analysis. Sociological Methods and Research, 1974, 2, 485-511.

Spearman, C. General intelligence, objectively determined and measured. American Journal of Psychology, 1904, 15, 201-293.

Thurstone, L. L. Multiple factor analysis. Psychological Review, 1931, 38, 406-427.

Tryon, R. C. Cluster analysis. Ann Arbor, Michigan: Edwards Brothers, Inc., 1939.

Tryon, R. C. Domain sampling formulation of cluster and factor analysis. Psychometrika, 1959, 24, 113-135.

Tryon, R. C., & Bailey, D. E. Cluster analysis. New York: McGraw-Hill Book Company, Inc., 1970.

Urry, V. W. Approximations to item parameters of mental test models and their uses. Educational and Psychological Measurement, 1974, 34, 253-269.

Appendix 1

The Use of PACKAGE and LISREL IV for Confirmatory Factor Analysis

Confirmatory Factor Analysis with PACKAGE

PACKAGE consists of a system of Fortran subprograms which includes (but is not limited to) programs for:

(a) exploratory guides for partitioning items into clusters: FACTOR (pp. 12, 14, and 16) and ORDER (p. 10).

(b) a multiple groups confirmatory factor analysis of a user-defined cluster solution: MGRP (pp. 12 and 16).

(c) statistical guides which aid in the evaluation of the goodness of fit of a cluster solution: PARTIAL (p. 25) and SQRR (p. 31).

The input to most PACKAGE subprograms is the square correlation matrix in core memory. One exception is the subprogram CORR (p. 12) which reads an original data matrix and computes item correlations. The output of most PACKAGE subprograms is another correlation matrix or a related matrix such as a matrix of similarity coefficients (p. 31). With a few exceptions, the output from one subprogram is available for input to a subsequent subprogram within the same run. However, execution of a PACKAGE subprogram usually replaces the matrix in core memory with the new matrix computed by the subprogram.

Utility routines are available for reading and writing correlation matrices. The correlation matrix which resides in core memory at any time during a run may be saved for later analyses by copying it to a local file. This matrix may be copied to the local file TAPE2 and saved as a PACKAGE binary system file with the routine COPYR, or it may be copied in BCD (decimal) form to the local file PUNCH with the routine PUNCH. Decimal

or binary forms of a correlation matrix may be read into core memory with the routine READR.

Data transformations are accomplished with a user-supplied Fortran subroutine called TRANS. However, the PACKAGE routine REFLECT allows selected items to be reflected, and routine REORDER reorders and/or deletes the correlations of items in the matrix residing in core memory.

The next three examples of PACKAGE programs illustrate the computation of the inter-item correlations and exploratory and confirmatory data analyses. The blank lines which separate subprograms or utility routines are included in these examples for heuristic purposes only. Explanations of the subprogram or utility routine are in italics. The formats for FACTOR and PUNCH are listed below. The formats for CORR, READR, and MGRP are listed in the following pages.

FACTOR format (all parameters optional)

COMM: The factor analysis will be done with the largest correlation for each variable as its communality. (The subsequent multiple groups analysis is always done with communalities.)
NF=m: The maximum number of factors is m. The maximum allowed and the default are 20.
RESID: The residual covariance matrix is printed.
RESID, LU=k: The residual covariance matrix is written on logical unit number k and is not printed.
NOCOMP: Do not print the principal components.

PUNCH format

no parameter specified: format is 5X,25F3.2
DIGITS: format is 10X,10F7.5 (optional)
SPSS: format is 8F10.7 (optional)

Example 1: The computation of the correlation matrix.

```
ATTACH,TAPE1,mydata.
ATTACH,PACKAGE,PACKAGE.
PACKAGE.                                    CDC control cards
CATALOG,TAPE2,mybinarycorrmatrix.
CATALOG,PUNCH,mydecimalcorrmatrix.
7/8/9

START,NMAX=20*                              {maximum number of variables, including
                                            {factors, is to be 20 on this run

CORR,NV=20,NS=351,NFC=1,LU=1,MD,BLANKS*
VARS(1-20)                                  {compute correlations
(8X,20F1)

REFLECT*                                    {reflect (i.e., reverse score) the
VARS(3,4,6,7,9,10,11,14,16,17)              {designated variables

COPYR,LU=2*                                 {copy the binary version of the
                                            {correlation matrix to local file TAPE2

PUNCH,SPSS*                                 {copy the decimal version of the correlation
                                            {matrix to local file PUNCH in SPSS format

END*                                        {end of program

6/7/8/9
```

CORR format

Required	NV:	number of variables
	NS:	number of subjects (may be larger than the actual number unless the data is read in from cards)
	NFC:	number of format cards
	LU:	logical unit number of the local file which contains the data, i.e., TAPE1. (If the data are on cards, use LU=60 and place the data after the format card.)
Optional	TPOSE:	transpose the data matrix before computing correlations
	MD:	missing data exist
	BLANKS:	missing data are coded as blanks.
	MD=k:	missing data are coded as the number k (e.g., MD=9)

The CORR card is followed by (a) a VARS card which lists the variables to be correlated in the order desired in the output, and (b) a format card which specifies the format of each case of the data with the F specification. The variables are numbered from 1 to NV as they are read in according to the FORMAT card.

Example 2: Exploratory factor analysis and an ordering of the correlation matrix.

```
ATTACH,TAPE1,mybinarycorrmatrix.
ATTACH,PACKAGE,PACKAGE.                    CDC control cards
PACKAGE.
7/8/9

START,NMAX=40*

READR,LU=1,NV=20,NOPRINT*
                                  {principal axis factors with communalities
FACTOR,COMM*                      {and a varimax rotation followed by a multiple
                                  {groups analysis of the resulting clusters
READR,LU=1,NV=20,NOPRINT*

ORDER*                            {order the correlation matrix

END*

6/7/8/9
```

READR format

The format of the READR statement is simpler for correlation matrices which are in PACKAGE binary form. The binary matrix can be attached as local file TAPE1 and read into core memory by READR with only the parameters LU and NV and, if desired, the NOPRINT option.

The format of the READR routine for a BCD matrix is only slightly more complicated:

```
READR,LU=1,NV=__,NFC=1,LABELS*
(format)
LABELS(numerical labels)
```

If there are 20 variables the LABELS card could look like LABELS(1-20) or perhaps LABELS(1-10,31-40). The format is the format of each card of the correlation matrix. For matrices obtained from the routine PUNCH, use (5X,25F3.2). For matrices computed by SPSS, use (8F10.7). The NOPRINT option may also be used in this form of the READR statement. If the correlation matrix is on cards, use LU=60 and place the matrix right after the format card.

Example 3: An oblique multiple groups analysis (and additional analyses) with PACKAGE

```
ATTACH,TAPE1,mybinarycorrmatrix.
ATTACH,PACKAGE,PACKAGE.                          CDC control cards
PACKAGE.
CATALOG,PUNCH,mydecimalfactorcorrmatrix.
7/8/9

START,NMAX=40*

READR,LU=1,NV=20,NOPRINT*

MGRP,NG=5,COMM*
DECEIT(7,6,9,10)                    { an oblique multiple groups confirmatory
FLATTERY(15,2)                      { factor analysis
IMMORALITY(11,16,4)
VICIOUS(12,5,13,1)
RESIDUAL(3,8,14,17-20)

COPYR,LU=2*                         { copy the correlation matrix, which now includes
                                    { communalities and the group factors, to TAPE2

REORDER*                            { include only the correlations among the group factors
VARS(501-504)                       { in the correlation matrix in core memory, i.e., drop
                                    { variables 1-20 and 505.

PUNCH*                              { obtain a BCD version of the group factor correlation
                                    { matrix for input to a path analysis program

READR,LU=2,NV=25,NOPRINT*           { read the correlation matrix from TAPE2, which
                                    { includes 20 items and 5 factors

PARTIAL*                            { partial out the group factors, which are numbered
VARS(501-504)                       { from 501 upwards (ignoring the residual group
                                    { factor 505)

READR,LU=1,NV=20,NOPRINT*
                                    { obtain similarity coefficients of the original 20
SQRR,DIAG=30*                       { items, set the diagonal of the matrix at .30 to
                                    { approximate communalities

END*

6/7/8/9
```

MGRP format

NG: number of groups (required);
group names and defining variables are supplied by the user on succeeding cards

COMM: communalities (optional)

Confirmatory Factor Analysis with LISREL IV

The following discussion is based on material contained in the LISREL IV manual by Jöreskog and Sörbom (1978).

Input. LISREL is a general program for the simultaneous parameter estimation and evaluation of measurement and causal models. There are, however, several ways to "trick" the program into analyzing only a measurement model and one of these ways will be illustrated below. In the LISREL IV program presented in Example 4, the numbers which are unique to this particular example are circled. The circled numbers have three values: 13 for the number of items, 4 for the number of factors, and 351 for the number of persons. In addition to replacing the circled numbers with the appropriate values, there are three other sets of changes: "Table A", "String A", and "String B".

The set of cards or lines labelled Table A in Example 4 represents, in a multiple groups analysis, the partitioning of the items into mutually exclusive clusters. These cards encompass several assumptions about the correlation matrix and the variables in the study. First of all, Table A has 13 rows of 4 entries corresponding to the 13 items and 4 factors of the example. Second, the pattern of 1's and 0's presumes a certain ordering of the variables in the input correlation matrix. The first 4 rows have a 1 in column 1 and a 0 in columns 2-4. This corresponds to the assumption that the first 4 items for the input correlation matrix are the indicators of factor 1. The next 2 rows have a 1 in column 2 and 0's elsewhere which correspond to the assumption that the 5th and 6th variables for the input correlation matrix are the indicators of factor 2. The next 3 rows have a 1 in column 3 and 0's elsewhere which specify that items 7 through 9 are the indicators of factor 3. The last 4 rows have a 1 in column 4 and 0's elsewhere indicating that variables 10 through 13 are linked to factor 4.

In general, Table A should have as many rows as items and as many columns as factors.

The rule for constructing "String A" is this: a zero, one 1, a zero, two 1's, a zero, three 1's, a zero, four 1's, a zero, ... The string is long enough when the number of zeros reaches the number of factors.

The rules for constructing "String B" are: (a) put a space between each pair of numbers and (b) start with a 1.0 followed by one .3, then a 1.0 followed by two .3's, then a 1.0 followed by three .3's, then a 1.0 followed by four .3's, ... The string is long enough when the number of 1.0's reaches the number of factors.

Output. The primary output of LISREL is (a) the set of matrices which contain the parameter estimates and (b) the χ^2 test of fit of the null hypothesis that the observed correlations conform to the structure implied by the factor model. These parameter matrices are:

LAMBDA Y (Λ_y): This estimated factor pattern matrix contains the regression coefficients of the observed variables regressed on those factors which appear in the model. Since each observed variable is "caused" by one only factor in a multiple groups analysis, there are non-zero regression coefficients only between each item and its corresponding factor. That is, the 0's in the table are specified by the model. The weight for each item is the estimated correlation between that item and its factor.

PSI (Ψ): The factor correlation matrix.

THETA EPS (Θ_ϵ): The item uniqueness or, in the language of this paper, the measurement error associated with each item. In a multiple groups analysis, this value for each item will be the square root of 1 minus the square of the correlation between that item and its factor.

Example 4: An oblique multiple groups analysis with LISREL IV

```
ATTACH,TAPE1,mydecimalcorrmatrix.    {first card of this file is the format card
HAL,L*UNSUP,L1B.
LISREL,PARM=DRTRTZBRXW.
7/8/9
400
FOUR FACTOR OBLIQUE MULTIPLE GROUPS ANALYSIS EXAMPLE          {title
DATAPARAMETERS NINPUTVARS=13 NOBS=351 MATRIX=KM
KM=FULL UNIT=1
MODELPARAMETERS NY=13 NETA=4 NX=0 NKSI=0  C
  BETA=IDENTITY PSY=SYMMETRIC,FREE TEPS=DIAGONAL,FREE LY=FULL,FIXED
PA LY
(4I1)
1000
1000
1000
1000
0100
0100
0010
0010                    Table A
0010
0001
0001
0001
0001
PA PSI
(80I1)
0101101110}              String A
START 0.5 LY(1,1)-LY(13,4)
START 0.8 TEPS(1)-TEPS(13)
MATRIX PSI
*
1.0 .3 1.0 .3 .3 1.0 .3 .3 .3 1.0}              String B
OUTPUTPARAMETERS PM MR FS SE PC TV FD SS

6/7/8/9
```

LABELS format

The observed variables in the above example are given numeric labels from 1 to 13 in the order they are read by the program. The user has the option, however, of supplying alphanumeric labels for these variables by inserting the following cards immediately after the "DATAPARAMETERS" card. There is one label for each observed variable. In the following example, the format specifies 10 labels per line, 6 characters per label (including spaces).

```
LABELS
(10A6)
DEC07 DEC06 DEC09 DEC10 FLT15 FLT02 IMM11 IMM16 IMM04 VIC12
VIC05 VIC13 VIC01
```

Additional matrices are printed by LISREL if these matrices are specified on the "OUTPUTPARAMETERS" card according to the two letter codes which are enclosed in parentheses below. For each model parameter, there is:

(a) a standard error (SE)

(b) a z value to test the null hypothesis that the corresponding parameter is zero (TV)

(c) the first derivative of the parameter (FD)

(d) the start value of the parameter estimate (PM)

If a code such as SE (for standard error) is on the "OUTPUTPARAMETERS" card, then three matrices are printed for that code which are also labelled "LAMBDA Y", "PSI", and "THETA EPS". The code PM also causes the sample inter-item correlation matrix to be printed.

The following matrices can also be obtained. The labels by which they are identified on the printout and the meaning of their contents are given below:

Y - ETA: The factor loading matrix, i.e., the matrix which contains the estimated correlations between the observed variables and all of the factors. (PC)

SIGMA: The correlations among the observed variables predicted by the model. For a multiple groups analysis, these correlations conform perfectly to the product rules for internal and external consistency. (MR)

RESIDUALS: The residuals defined by the subtraction of the predicted correlations from the observed correlations. (MR)

ETA ON Y: The regression coefficients for computing the factor scores from the observed variables. The program does not compute the scores themselves.

More details for LISREL users. LISREL maintains the distinction between a)

true scores or factors and b) observed scores or items or indicators of the true scores. Since causal models are usually included in LISREL analyses, the distinction is also made between endogenous and exogenous variables. Thus there are four kinds of variables in LISREL described by the following notation:

	Observed	Unobserved
Exogenous	x	ξ
Endogenous	y	η

The covariance matrices of the ξ's and η's are Φ and C respectively. There are two kinds of error variables: (a) errors of measurement, denoted δ for each observed exogenous variable and ϵ for each observed endogenous variable, and (b) errors of prediction for each regression equation between factors, denoted ζ with covariance matrix Ψ. Regression coefficients defined by regressing endogenous variables on exogenous variables are called γ's and the regression coefficients defined by regressing endogenous variables on each other are called β's. The matrices which contain these coefficients are, respectively, Γ and B.

The program will run only if (a) endogenous variables are included in the model, and (b) causes of the endogenous variables, either ξ's or ζ's or both, are specified. The underlying factors are conceptualized as endogenous variables in this example. The program is run without exogenous variables (NX=0 and NKSI = 0), but the ζ's are included. The observed variables are defined by the y matrix and the factors, i.e., the latent variables, are defined by the η matrix. The problem is that the correlations among η's can be specified only through causal paths, i.e., via γ's and β's, and the correlations among the factors in a confirmatory factor analysis are not explained in terms of causal paths. But there is a

solution.

Since there are no exogenous variables, the model of η

$$B\eta = \Gamma\xi + \zeta$$

becomes

$$B\eta = \zeta$$

Since no causal structure among the η's is specified, the B matrix is set to an identity matrix. That is, the model of η reduces to:

$$\eta = \zeta$$

So although the user cannot directly specify the correlations among the η's, it is possible to do so with the Ψ matrix, which is the covariance matrix of the ζ's.

The matrix form of the model of interest, the measurement model for endogenous variables, is:

$$y = \Lambda_y \eta + \varepsilon$$

The equation for any single y for a multiple groups solution is:

$$y_i = \lambda_{y_i} \eta_j + \epsilon_i$$

where y_i is an indicator of η_j. This formula is equivalent to the true score model presented on page 18 for $y_i = x_i$, $\lambda_{y_i} = \beta_i$, $\eta_j = T$, and $\epsilon_i = e_i$.

Since $\zeta = \eta$ for this application,

$$y = \Lambda_y \zeta + \Theta_\epsilon$$

with covariance matrix:

$$\Sigma_{yy} = \Lambda_y \Psi \Lambda'_y + \Theta_\epsilon$$

So the parameters to be estimated are defined in the following three matrices:

Λ_y: factor pattern

Ψ : factor-factor correlation

Θ_ϵ : item uniqueness

Traditionally, a factor analysis is based on a correlation matrix. The correlation among the factors is obtained by specifying the diagonal of the Ψ matrix to be fixed with 1's. The remainder of Ψ is free, as is the Θ_ϵ or measurement error matrix. Since there is no specified causal structure among the factors, the B matrix is defined as the identify matrix. The Λ_y or factor pattern matrix is fixed at zero except for the indicators of each factor which are free. Since NXVAR = 0, the Λ_x, Γ, Φ, and Θ_δ matrices are not defined.

The LISREL IV program for confirmatory factor analysis presented in Example 4 illustrates the most common application: an oblique multiple groups analysis. That is, (a) each observed variable is an indicator of only one factor (each row of the pattern matrix for Λ_y contains only one fixed element), and (b) the correlations among the factors are not constrained to be zero (the off-diagonals of Ψ are free). However, LISREL can also be used to test general factor models (as in Appendix 4) in which each observed variable may be defined by multiple factors.

Appendix 2

The Computation of a PACKAGE "Oblique Multiple Groups Factor Analysis"

The computations of a factor analysis are the computations of parameter estimates. These computations used by PACKAGE will be explained for the three kinds of parameters in a multiple groups analysis: factor loadings, factor-factor correlations, and communalities.

Computation of the Factor Loadings

The factor loading of item X onto factor F is the correlation of the item and the factor, which is by definition:

$$r_{XF} = \frac{\sigma_{XF}}{\sigma_X \sigma_F}$$

The computation of a factor loading follows from the computation of the item-factor covariance and the item and factor standard deviations.

The computation of the item-factor covariances is suggested by a key fact of the classical multiple groups analysis: each factor is a centroid factor extracted from the corresponding group. That is, in a solution without communalities, a factor score is simply defined as the sum of the responses to the items which define the group or cluster. So if the group factor F is defined by items Y_1 through Y_n, then, by the usual formula for the covariance of a sum (e.g., Nunnally, 1978, p. 163):

$$\sigma_{XF} = \sigma_{X,\ Y_1 + Y_2 + \ldots + Y_n}$$

$$= \sigma_{XY_1} + \sigma_{XY_2} + \ldots + \sigma_{XY_n}$$

The covariance of an item with a group factor is simply the sum of the covariances of the item with each of the items which define the corresponding group. If the variables are standardized, then the estimates of these item covariances are simply the observed corresponding item

correlations, i.e.,

$$\sigma_{XF} = r_{XY_1} + r_{XY_2} + \ldots + r_{XY_n} \quad \text{if} \quad \sigma_X = \sigma_{Y_i} = 1$$

All we need to complete the formula for a factor loading are the standard deviation of the item X and the factor F.

The standard deviation of the item is 1 for standardized variables. The computation of the factor standard deviation is given in the next section.

Computation of Factor Correlations

By definition, the formula for the correlation between factors F_i and F_j is:

$$r_{F_iF_j} = \frac{\sigma_{F_iF_j}}{\sigma_{F_i}\sigma_{F_j}}$$

Since each factor score is simply the sum of the item scores which define the factor, we have:

$$F_i = X_1 + X_2 + \ldots + X_n$$
$$F_j = Y_1 + Y_2 + \ldots + Y_m$$

where X_1 through X_n are the n items defining the group factor F_i and Y_1 through Y_m are the m items defining group factor F_j. Then the covariance between the factors is, by the formula for the covariance of two sums:

$$\sigma_{F_iF_j} = \sigma_{X_1Y_1} + \sigma_{X_1Y_2} + \ldots + \sigma_{X_1Y_m} + \sigma_{X_2Y_1} + \ldots + \sigma_{X_nY_m}$$

The covariance between two group factors is the sum of all the covariances between the items which define the factors.

However, another version of the formula for factor covariances is used in the computation:

$$\sigma_{F_iF_j} = \sigma_{F_iY_1} + \sigma_{F_iY_2} + \ldots + \sigma_{F_iY_m}$$

This formula is used since these item-factor covariances have already been computed for use in the computations of the factor loadings.

What about the standard deviation of the factors which we need to complete the computations of both the item-factor correlations and the factor-factor correlations? The key here is that a standard deviation of a variable is the square root of the variable's covariance with itself. That is:

$$\sigma_{F_j} = \sqrt{\sigma_{F_j}^2} = \sqrt{\sigma_{F_j F_j}}$$
$$= \sqrt{\sigma_{F_j Y_1} + \sigma_{F_j Y_2} + \ldots + \sigma_{F_j Y_m}}$$

The standard deviation of a group factor is equal to the square root of the sum of the covariances of the factor with each of the items which define the factor.

Computation of the Estimated Communalities

All of the computational formulas for the factor loadings and factor correlations for a multiple groups analysis of centroid factors have been presented. Any partitioned correlation matrix of observed item scores could be factored using these formulas. The factoring of a covariance matrix of true scores is accomplished by inserting true score variances or communalities in the diagonal of the original correlation matrix (see Appendix 3). No formula exists for the computation of these communalities, but they can be estimated using an iterative process based on the formulas for factor loadings and factor correlations already developed. This iterative procedure generates communalities for each cluster one at a time. It works only if the primary assumption of the measurement model is met: all of the items in that cluster have the same true score. If the postulated unidimensional measurement model is not correct, then the

estimated communalities for that cluster are not correct.

The key to the iterative procedure is the formula:

$$r_{X_i X_i} = r^2_{X_i T}$$

That is, the communality of an item is equal to the squared factor loading of the item on its own group factor (see page 20). Thus if the correct number is placed in the diagonal, then the resulting part whole correlation calculated using that number will be the square of the number in the diagonal. If an incorrect number is inserted in the diagonal, then it can be shown that the square of the resulting part whole correlation will be closer to the desired communality by an amount proportional to the discrepancy. The constant of proportionality is an increasing function of cluster size.

The analysis begins with 1's in the diagonal. The resultant factor loadings are squared to provide an estimate of the corresponding communalities. These estimates are not correct, but they are closer to the correct values (based on the sample data) than the original 1's with which the analysis began. The analysis is repeated with the revised communality estimates and a new set of factor loadings, and hence communalities, are obtained. This procedure yields successively better estimates, though usually in less than five iterations the successive communality estimates for each item are equal to within two decimal places.

Appendix 3

The Meaning of Communalities

There are two kinds of parameters in a measurement model: item-factor correlations and factor-factor correlations. If the correlation matrix is factored without communalities, the factors are simply the sums of the observed responses to the items which define each cluster. The goal, however, is to define the factors to be cluster true scores so that the factor loadings are the estimated correlations between items and cluster true scores, and the factor-factor correlations are estimated correlations between cluster true scores.

The correlation matrix of observed item scores is given. The problem is: (a) How can this matrix be transformed into a covariance matrix for item true scores? and (b) What are the characteristics of such a matrix? First consider the observed correlations between different items. These correlations are the covariances among the standardized item observed scores, i.e., $r_{X_iX_j} = \sigma_{X_iX_j}$. These covariances are due only to true scores since the error of one item does not correlate with either the true score or error of another item.

To show this mathematically, let $X_i = T_i + e_i$ and $X_j = T_j + e_j$. The covariance between items X_i and X_j for $i \neq j$ may be expressed as:

$$\begin{aligned}\sigma_{X_iX_j} &= \sigma_{T_i + e_i, T_j + e_j} \\ &= \sigma_{T_iT_j} + \sigma_{T_ie_j} + \sigma_{T_je_i} + \sigma_{e_ie_j} \\ &= \sigma_{T_iT_j}\end{aligned}$$

That is, for $i \neq j$, $\sigma_{T_iT_j} = \sigma_{X_iX_j}$. So the covariance of the item true scores is equal to the correlation between the item observed scores

for standardized item scores.

What are the item true score variances? The variance of the standardized observed scores is 1.00, which appears in the diagonal of the correlation matrix of observed scores. The variance of an observed item is composed of both true and error variance, i.e.:

$$\sigma_{X_i X_i} = \sigma^2_{X_i} = 1 = \sigma^2_{T_i} + \sigma^2_{e_i} > \sigma^2_{T_i}$$

The variance of observed scores is not equal to the variance of true scores, but is greater by the size of the error variance.

$$\sigma^2_{T_i} = \sigma^2_{X_i} - \sigma^2_{e_i} = 1 - \sigma^2_{e_i}$$

This number, which is less than

replaces the original 1 in the diagonal. So the communality of item X_i is equal to its true score variance.

The off-diagonal elements of the given correlation matrix of observed item scores are also the item true score covariances. The communalities of the items are the item true score variances. Thus the insertion of communalities in the observed score correlation matrix transforms this matrix into a true score covariance matrix.

Appendix 4

The General and Unidimensional Factor Models

The discussion of unidimensionality in this paper has generally been phrased in the language of reliability theory and measurement. But, except for differences in notation, these "measurement models" could also be called "factor models". The unidimensional factor model, which was conceived by Spearman in 1904, is an example of the more general factor model introduced by Thurstone in 1931.

Spearman's unidimensional factor model of an individual's response to the j^{th} item is:

$$X_j = \beta_j F + e_j$$

This is the same model presented in page 18 for F = T. Since Spearman's method was only applicable to unidimensional variable sets, researchers searched for a technique which would provide for the analysis of multi-dimensional data. The "multiple groups factor analysis" used in this paper is a generalization of Spearman's model to multiple variable sets (or "groups") where each variable set represents a different dimension.

Thurstone's alternative solution was to simultaneously extract several orthogonal (i.e., uncorrelated) factors in what was called "multiple factor analysis". Thurstone's method evaded the issue of content analysis by using purely mathematical rules for defining factors. Refinement of these methods has led to today's exploratory factor analysis: "principal axis factor analysis" (or some variant) followed by "rotation". The elimination of content analysis led to simple computer programs which were easily and widely promulgated. In fact, for 25 years the term "factor analysis" was generally used synonymously with exploratory factor analysis.

However, evading the issue of content analysis at the beginning just postpones the problem to the end where it is called "interpreting the factors". The problem in this form proved so hard that books on factor analysis written during this period have very little to say on the issue!

The general factor analysis model (e.g., Gorsuch, 1974; Harmon, 1976; and Nunnally, 1978) of an individual's response to the j^{th} item is:

$$X_j = \beta_{1h}F_1 + \beta_{2j}F_2 + \ldots + \beta_{mj}F_m + S_j + e_j$$

There are m common factors, i.e., m factors which at least two items on the scale correlate with. There is one factor, S_j, which is specific to X and a random error component, e_j. Both S_j and e_j are uncorrelated with any other items in the scale, but the specific factor S_j may be correlated with variables not included in the particular scale. In terms of reliability, item specific variance is true variance, as is the variance contributed by the common factors. In terms of factor analysis, item specific variance "acts like" random error variance. Thus, in the application of the model, it is impossible to distinguish between the specific and random error components from the analysis of any one scale at one point in time. A new term, E_j, is often formed according to:

$$E_j = S_j + e_j$$

where E_j is called the "item uniqueness". This "composite" error term is the error component of the unidimensional factor model.

The "fundamental theorem of factor analysis" states that if the factors are uncorrelated with each other, then the correlation between two items can be reproduced as a sum of the products of the factor loadings across all the common factors. (The corresponding rule for correlated factors is more complex.) Since S_j and e_j are uncorrelated with all other items, they do not contribute to the correlation of X_j with any other X_k.

$$r_{X_j X_k} = \beta_{1j}\beta_{1k} + \beta_{2j}\beta_{2k} + \ldots + \beta_{mj}\beta_{mk}$$

This general factor theorem is the multiple factor generalization of the "product rule for internal consistency" which is applicable to the one factor model. That is:

$$r_{X_j X_k} = \beta_{1j}\beta_{1k}$$

where there is only one common factor.

The communality of item X_j, represented by h_j^2, is the variance the item has in common with the remaining items as accounted for by the m factors, i.e.:

$$h_j^2 = R^2_{X_j \cdot F_1 F_2 \ldots F_m}$$

For uncorrelated factors, the communality of item X_j can be computed by applying the fundamental theorem with k = j.

$$h_j^2 = \beta_{1j}^2 + \beta_{2j}^2 + \ldots + \beta_{mj}^2$$

This formula reduces to:

$$h_j^2 = \beta_j^2$$

for the unidimensional model. The communality formula above was often interpreted as a "decomposition of the common variance into its factorial components". However, for correlated factors, the formula is more complex and cannot be interpreted as a "decomposition." The conceptually meaningful factors in most substantive areas are correlated with each other and hence the uncorrelated factors (and simple formulas) of an exploratory orthogonal factor analysis have no practical significance in their own right.

Appendix 5

Proof of the Product Rules for Internal and External Consistency

Assumption

Assume the classical true score model in which the observed score has a linear regression on the true score. That is:

$X = \beta T + e$ where $E(X|T) = \beta T$ and hence $r_{e,T} = 0$

By the usual formula for linear regression we have:

$$\beta = \frac{\sigma_{XT}}{\sigma_T^2} \quad \text{or} \quad \beta = r_{XT} \text{ if X and T are standardized}$$

If X and T are standardized, e is usually not converted to standard scores, but is left to have variance:

$$\sigma_e^2 = 1 - r_{XT}^2$$

If the traditional $X = T + e$ model is used, then T is not standardized either.

Internal Consistency

Let X_i and X_j be measures of the same latent variable. Then if the errors e_i and e_j are uncorrelated,

$$r_{X_iX_j} = r_{X_iT}\, r_{X_jT}$$

Proof: By definition, $r_{X_iX_j} = \frac{\sigma_{X_iX_j}}{\sigma_{X_i}\,\sigma_{X_j}}$, so the first task is to find $\sigma_{X_iX_j}$. Since $X_i = \beta_i T + e_i$ and $X_j = \beta_j T + e_j$,

$$\begin{aligned}
\therefore \sigma_{X_iX_j} &= \sigma_{\beta_iT + e_i,\ \beta_jT + e_j} \\
&= \sigma_{\beta_iT,\beta_jT} + \sigma_{\beta_iT,e_j} + \sigma_{e_i,\beta_jT} + \sigma_{e_i,e_j} \\
&= \beta_i\beta_j\sigma_{TT} + \beta_i\sigma_{Te_j} + \beta_i\sigma_{e_iT} + \sigma_{e_ie_j} \\
&= \beta_i\beta_j\sigma_{TT} \\
&= \beta_i\beta_j\sigma_T^2
\end{aligned}$$

If X_i, X_j, and T are in standard scores, then $\sigma_{X_iX_j} = r_{X_iX_j}$, $\beta_i = r_{X_iT}$ and $\sigma_T^2 = 1$, so we immediately have:

$$r_{X_iX_j} = r_{X_iT}\, r_{X_jT}$$

If X_i, and X_j, and T are not standardized;

$$r_{X_iX_j} = \frac{\sigma_{X_iX_j}}{\sigma_{X_i}\,\sigma_{X_j}} = \frac{\beta_i\,\beta_j\,\sigma_T^2}{\sigma_{X_i}\,\sigma_{X_j}} = \frac{\beta_i\,\sigma_T}{\sigma_{X_i}} \quad \frac{\beta_j\,\sigma_T}{\sigma_{X_j}}$$

But $r_{X_iT} = \frac{\sigma_{X_iT}}{\sigma_{X_i}\,\sigma_T}$ and $\beta_i = \frac{\sigma_{X_iT}}{\sigma_X^2}$, so $r_{X_iT} = \frac{\beta_i\,\sigma_T}{\sigma_{X_i}}$

By substitution,

$$r_{X_iX_j} = r_{X_iT}\, r_{X_jT}$$

External Consistency

Case 1: Let X be a measure of T with error e, and let U be any other variable.

If e, the error of measurement in X, is uncorrelated with U, then:

$$r_{XU} = r_{XT}r_{TU}$$

Proof: By definition, $r_{XU} = \frac{\sigma_{XU}}{\sigma_X\,\sigma_U}$, so the first task is to find σ_{XU}.

Since $X = \beta T + e$,

$$\sigma_{XU} = \sigma_{\beta T + e,\, U}$$

$$= \beta\,\sigma_{TU} + \sigma_{eU}$$

$$= \beta\,\sigma_{TU}$$

If the variables X, T, and U are in standard score form, then:

$$\sigma_{XU} = r_{XU}, \quad \beta = r_{XT}, \text{ and } \sigma_{TU} = r_{XU}$$

so: $r_{XU} = r_{XT}r_{XU}$

and we are done.

In the general case:

$$r_{XU} = \frac{\sigma_{XU}}{\sigma_X \sigma_U} = \frac{\beta \sigma_{TU}}{\sigma_X \sigma_U} = \frac{\beta}{\sigma_X} \cdot \frac{\sigma_{TU}}{\sigma_U} = \frac{\beta}{\sigma_X} \cdot \frac{\sigma_T}{\sigma_T} \cdot \frac{\sigma_{TU}}{\sigma_U}$$

$$= \frac{\beta \sigma_T}{\sigma_X} \cdot \frac{\sigma_{TU}}{\sigma_T \sigma_U}$$

$$= r_{XT} \cdot r_{TY}$$

Case 2: Let X be a measure of T and let Y be a measure of any other latent variable U.

If the errors of measurement are uncorrelated with each other and with the latent traits, then:

$$r_{XY} = r_{XT}r_{TU}r_{UY}$$

Proof: From Case 1, if Y = U, then:

$$r_{XY} = r_{XT}r_{TY}$$

Also from Case 1, r_{TY} can be expanded according to:

$$r_{TY} = r_{YU}r_{UT}$$

By substitution,

$$r_{XY} = r_{XT}r_{TU}r_{UY}$$

PUBLICATIONS

of the

Institute for Research on Teaching
Michigan State University

as of

February 1, 1979

To order any of the following publications please send check, money order, or prepaid purchase order -- payable to Michigan State University -- to: IRT Publications, 252 Erickson, MSU, East Lansing, MI 48824. Publication prices include only the cost of production and mailing. Michigan residents should add a 4% state sales tax to all orders. The Institute for Research on Teaching is funded primarily by the National Institute of Education, United States Department of Health, Education, and Welfare.

Research Series

No. 1 The Inquiry Theory: An information-processing approach to clinical problem-solving research and application. J.F. Vinsonhaler, C.C. Wagner, & A.S. Elstein. 1978. $2.00

No. 2 Impact on what? The importance of content covered. A.C. Porter, W.H. Schmidt, R.E. Floden, & D.J. Freeman. 1978. $1.75

No. 3 Instructions for using the AJ 832 plotting software package as modified for the CDC 6500 computer. M. Carlyn. 1977. $1.75

No. 4 Training manual for the classification of the content of fourth-grade mathematics. W.H. Schmidt, A.C. Porter, R.E. Floden, & D.J. Freeman. 1978. $2.25

No. 5 Flexner, accreditation, and evaluation. R.E. Floden. 1978. $1.75

No. 6 Analogy and credentialling. R.E. Floden. 1978. $1.25

No. 7 Conceptions of reading: The Rep Test. M. Johnston. 1978. $1.00

No. 8 Research and development needs for the advancement of teacher education. J.E. Lanier & R.E. Floden. 1978. $3.00

No. 9 On the conceptualization of clinical problem solving. C.C. Wagner & J.F. Vinsonhaler. 1978. $1.50

* No. 10 Applications of the inquiry theory to reading and learning disabilities. L. Patriarca, J. VanRoekel, & J.F. Vinsonhaler. 1979.

No. 11 Methods for discovering cues used by judges: Two working papers. C.M. Clark & R.J. Yinger. 1978. $1.00

No. 12 Research on teacher thinking. C.M. Clark & R.J. Yinger. 1978. $2.25

No. 13 Data analysis strategies for quasi-experimental studies where differential group and individual growth rates are assumed. S. Olejnik. 1978. $2.75

No. 14 CLIPIR Pilot Observational Study of Reading Diagnosticians, 1976. A. Lee & A. Weinshank. 1978. $2.00 (with complete data, $5.75)

No. 15 The Inquiry Theoretic approach to clinical problem-solving research and application: Basic Management Information System (BMIS) technical manual Vol. 1. J.F. Vinsonhaler, C.C. Wagner, A.S. Elstein, & L.S. Shulman. 1978. $4.50

No. 16 The user's manual for the Basic Management Information System: BMIS technical manual Vol. 2. C.C. Wagner, J.F. Vinsonhaler, A.S. Elstein, & L.S. Shulman. 1978. $17.00

No. 17 Teachers' conceptions of reading: The evolution of a research study. R. Barr & G.G. Duffy. 1978. $1.25

No. 18 A study of teacher planning: Description and model of preactive decision making. R.J. Yinger. 1978. $4.25

No. 19 Fieldwork as basis for theory building in research on teaching. R.J. Yinger. 1978. $2.25

No. 20 Choice of a model for research on teacher thinking. C.M. Clark. 1978. $1.50

No. 21 Conceptual issues in the content/strategy distinction. D.J. Freeman. 1978. $2.25

No. 22 An econometric perspective on classroom reading instruction. B.W. Brown & D.H. Saks. 1978. $1.75

No. 23 Identifying cues for use in studies of teacher judgment. C.M. Clark, R.J. Yinger, & S.C. Wildfong. 1978. $1.75

No. 24 Teacher autonomy and the control of content taught. A.C. Porter. 1978. $3.75

No. 25 Don't they all measure the same thing? Consequences of standardized test selection. R.E. Floden, A.C. Porter, W.H. Schmidt, & D.J. Freeman. 1978. $1.50

* No. 26 Critical moments in the teaching of mathematics. J.C. Shroyer. 1978.

No. 27 Characteristics of the clinical problem-solving model and its relevance to educational research. (Formerly available as an IRT collateral paper.) A.S. Elstein, L.S. Shulman, J.F. Vinsonhaler, C.C. Wagner, & L. Bader. 1978. $1.25

* No. 28 The consistency of reading diagnosis. A. Weinshank. 1978.

* No. 29 The Inquiry Theory research paradigm. C.C. Wagner. 1978.

* No. 30 Simulating the problem solving of reading clinicians. J.F. Vinsonhaler. 1978.

* No. 31 Training reading specialists in diagnosis. G. Sherman. 1978.

No. 32 Classroom management in the elementary grades. J.E. Brophy & J.C. Putnam. 1978. $3.25

No. 33 An ethnographic study of a teacher's classroom perspective: Implications for curriculum. V.J. Janesick. 1978. $2.25

No. 34 The problem of dead letters: Social perspectives on the teaching of writing. S. Florio. 1978. $1.50

No. 35 Measuring content of instruction. W.H. Schmidt. 1978. $1.50

* No. 36 The relationship of teacher alienation to school workplace characteristics and career stages of teachers. M.J. Vavrus. 1978. $2.50

No. 37 The relationship between diagnosis and remediation in reading: A pilot study. A. Weinshank. 1978. $1.75

* No. 38 Teacher judgment of children's reading preferences. T.E. Evans & J.L. Byers. 1979. $2.00

* No. 39 The teaching styles at South Bay School. B. Joyce. 1979. $2.25

* No. 40 Teachers' thoughts while teaching. B. Joyce. 1979. $2.25

* No. 41 Measuring teachers' conceptions about reading. G.G. Duffy & W. Metheny. 1979. $1.75

* No. 42 Studies of clinical problem-solving behavior in reading diagnosis. D. Gil, J.F. Vinsonhaler, & C.C. Wagner. 1979. $2.00

* No. 43 Study of the value/belief patterns of teachers and administrators. P. Cusick. 1979.

No. 44 Teacher perceptions of student affect. R.S. Prawat. 1979. $2.50

Occasional Papers

No. 1 Teachers' concerns and conceptions of reading and the teaching of reading: A literature review. G. Belli, G. Blom & A. Reiser. 1977. Limited number of free copies.

No. 2 Teachers and researchers: Toward a proper division of labor. C. Kennedy. 1977. Limited number of free copies.

No. 3 A causal analysis of attitudes toward leadership training in a classroom setting. J.E. Hunter, R.F. Hunter & J.E. Lopis. 1978. $2.00

* No. 4 The teacher as colleague in classroom research. S. Florio & M. Walsh. 1978.

No. 5 Form and function in mother-toddler conversational turn-taking. M.L. Donahue. 1978. $1.75

No. 6 Individual school buildings do account for differences in measured pupil performance. L.W. Lezotte & J. Passalacqua. 1978. $1.25

No. 7 Research on teaching: A dynamic area of inquiry. J.E. Lanier. 1978. $1.25

No. 8 Test design: A view from practice. L.S. Shulman. 1978. $1.75

No. 9 Relationships between testing and curriculum. A.C. Porter. 1978. $1.00

No. 10 Psychology and mathematics education revisited: 1976. L.S. Shulman. 1978. $1.75

No. 11 Science and mathematics education: Retrospect and prospect. L.S. Shulman & P. Tamir. 1978. $2.00

No. 12 Relating theory to practice in educational research: A working paper. L.S. Shulman. 1978. $1.75

No. 13 Classroom discipline: Toward a diagnostic model integrating teachers' thoughts and actions. D. Gil & P.S. Heller. 1978. $1.00

No. 14 Attention and modality effects in STM: A second look. T.Evans & J. Byers. 1978. $1.75

* No. 15 Mere ethnography: Some problems in its use in educational practice. F. Erickson. 1979. $2.00

* No. 16 On standards of descriptive validity in studies of classroom activity. F. Erickson. $2.00

* No. 17 Changes in school characteristics coincident with changes in student achievement. W.B. Brookover, & L.W. Lezotte. 1979. $5.00 (Executive Summary $1.00)

Conference Series

No. 1 Current directions in research on teaching: A meeting of the Invisible College of Researchers on Teaching, November 17-19, 1976. 1977. $4.25

No. 2 Report of a seminar on field research methods in education. P.A. Cusick. 1978. $1.50

No. 3 Proceedings of the Research-on-Teaching Mathematics Conference, May 1-4, 1977. 1978. $8.25

No. 4 Teachers attaining new roles in research: A challenge for the education community. L.D. Shalaway, J.E. Lanier et al. 1978. $3.50

---

*In production and will be available by May 1979. Please write for exact price and publication date.